EPJ manuscript No.
(will be inserted by the editor)

Coverage centralities for temporal networks

Taro Takaguchi[1,2,ab], Yosuke Yano[2,3,a], and Yuichi Yoshida[1,4,a]

[1] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
[2] JST, ERATO, Kawarabayashi Large Graph Project, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
[3] Department of Computer Science, The University of Tokyo, 3-7-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
[4] Preferred Infrastructure, Inc., 2-40-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

the date of receipt and acceptance should be inserted later

Abstract. Structure of real networked systems, such as social relationship, can be modeled as temporal networks in which each edge appears only at the prescribed time. Understanding the structure of temporal networks requires quantifying the importance of a temporal vertex, which is a pair of vertex index and time. In this paper, we define two centrality measures of a temporal vertex based on the fastest temporal paths which use the temporal vertex. The definition is free from parameters and robust against the change in time scale on which we focus. In addition, we can efficiently compute these centrality values for all temporal vertices. Using the two centrality measures, we reveal that distributions of these centrality values of real-world temporal networks are heterogeneous. For various datasets, we also demonstrate that a majority of the highly central temporal vertices are located within a narrow time window around a particular time. In other words, there is a bottleneck time at which most information sent in the temporal network passes through a small number of temporal vertices, which suggests an important role of these temporal vertices in spreading phenomena.

PACS. 89.75.Fb Structures and organization in complex systems – 89.75.Hc Networks and genealogical trees – 64.60.aq Networks

1 Introduction

Complex networks such as social networks, information networks, and biological networks have been intensively studied in the past decade to understand their behavior under certain dynamics and develop efficient algorithms for them. See [1–4] for extensive surveys.

However, many real-world networks are actually temporal networks [5, 6], in which a vertex communicates with another vertex at specific time over finite duration. For example, social interaction between individuals, passenger flow between cities, and synaptic transmission between neurons can be represented as temporal networks. When we assume that the focal dynamical processes on networks, such as information propagation, occur on a time scale comparable to the change in network structure, a temporal-network representation gives us a precise way to capture the processes. We can describe the advantage of working with a temporal network using the example shown in Fig. 1. This temporal network consists of four vertices and eight edges, each of which has the time it appears. Let us assume that it takes unit time to send the information from the tail to the head of an edge. For example, suppose that the information starts to propagate from v_1 at time 1. Then, it reaches v_2 at time 2 through edge (v_1, v_2), waits at v_2 till time 3, then reaches v_3 at time 4 through edge (v_2, v_3). The information never reaches v_4 because the only edge incoming to v_4 is (v_2, v_4) which appears at time 1, and v_2 does not have the information at that time. However, if we ignore the temporal information and regard the network as a static directed network, we mistakenly reach the conclusion that information in v_1 at time 1 can reach v_4 because there is a directed path from v_1 to v_4. Therefore, we cannot dismiss temporal information to properly understand the structure of temporal networks.

An important notion studied to understand the structure of (static) networks is vertex centrality, which measures the importance of a vertex. The following reasons motivate the study of centralities. First, we can use centralities to find important vertices in several applications such as suppressing the epidemics [7, 8] or maximizing the spread of influence [9]. Second, we can use them to understand the structure of real-world networks by examining the difference between the distributions of the centrality values in such networks and in the randomized networks (e.g, [10, 11]). Third, we can examine the validity of generative network models by investigating the distribution of centralities of the generated network (e.g., [12, 13]).

Hence, it is natural to study centralities for temporal networks. Since the most fundamental difference between

[a] All the authors contributed equally to the work.
[b] *e-mail:* t_takaguchi@nii.ac.jp

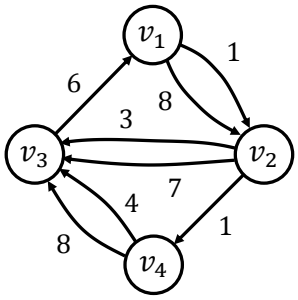

Fig. 1. Schematic of an example of temporal network. The number associated with each edge represents the time at which the edge appears.

a static network and a temporal network is that the latter involves time, we define the centrality of a vertex at a specific time. To distinguish from a vertex, we call the pair of a vertex and time a temporal vertex. In the literature, multiple centrality notions of temporal vertices based on temporal paths [5] have been proposed. Examples include the generalizations of the centrality notions to temporal networks, such as betweenness [15–18], closeness [16, 17, 19], communicability [20–22], efficiency [14], random-walk centrality [23], and win–lose score [24] (see Ref. [25] for a review of some of them). However, each previous centrality notion suffers from at least one of the following two issues:

1. We need to carefully set parameter values and (or) the time interval within which we consider temporal paths.
2. It is inefficient to compute the centrality.

For the first issue, the time interval length especially requires careful tuning; if the time interval is too wide, then the centrality of a temporal vertex $\boldsymbol{v}$ becomes negligible because most of the paths finish before or start after $\boldsymbol{v}$ appears. By contrast, if the time interval is too narrow, again the centrality of $\boldsymbol{v}$ becomes negligible because paths can pass by only a tiny fraction of vertices in the time interval. It should be noted that our centrality measures are free from any parameters not because we consider the centrality of temporal vertex. The centrality measures of a temporal vertex proposed in the previous work [14–25] require some parameters for different reasons. Our centrality measures get around this issue by focusing on the local structure of temporal paths around the focal temporal vertex. For the second issue, even if we compromise to use an approximation, computing the approximated centrality value of a single temporal vertex requires computational time at least linear to network size [26].

In this paper, we propose two novel centrality notions for temporal networks that resolve these issues. The first one, called temporal coverage centrality (TCC), measures the fraction of pairs of (normal) vertices that have at least one fastest temporal path that uses the focal temporal vertex. The second one, called temporal boundary coverage centrality (TBCC), measures the fraction of pairs of vertices that have a unique fastest temporal path, which uses the focal temporal vertex.

Our centrality notions address the two issues described above in the following way. For the first issue, TCC and TBCC are free from setting of any parameters or time interval. To calculate the TCC or TBCC value of a temporal vertex $\boldsymbol{v} = (v, \tau)$, we only have to run over all pairs of vertices (u, w). Namely, we consider temporal vertices $\boldsymbol{u} = (u, \tau_u)$ and $\boldsymbol{w} = (w, \tau_w)$, where τ_u is the latest time at which we can send information from u so that it reaches v at time τ, and τ_w is the earliest time at which we can receive information at w that is sent from v at time τ. It should be noted that, if we fix focal temporal vertex $\boldsymbol{v}$, τ_u and τ_w are uniquely determined by u and w, respectively, and that we thus do not have to care about the time interval around $\boldsymbol{v}$. Then, we check whether the information sent from $\boldsymbol{u} = (u, \tau_u)$ to $\boldsymbol{w} = (w, \tau_w)$ can or should drop by $\boldsymbol{v}$.

For the second issue, although the definitions of TCC and TBCC might look complicated and hard to compute, this is not the case. Indeed, computing TCC and TBCC can be reduced to the problem of deciding whether or not there is a directed path between queried vertices in an associated directed network (see Section 2.2 for details). The latter problem is well studied in the database community [38–42], and it can be solved by constructing an index of the directed network, which computes the reachability between any pair of nodes by using information of the reachability between a fraction of node pairs. If it suffices to use approximations to the TCC and TBCC values, we only need to query the index at most $\mathrm{O}(\log^2 N)$ times, where N is the total number of vertices in the network (see Appendix A). Since we can efficiently process queries to the index in practice, this method is advantageous compared to the $\mathrm{O}(N)$ time for approximating previous centrality notions.

With the aid of our centrality notions, we are able to compute the centrality of all temporal vertices in a temporal network and analyze the statistics of the whole network. Using TBCC, we demonstrate that real-world temporal networks have a small number of temporal vertices without which information propagates more slowly. Surprisingly, we reveal that the temporal vertices of large centrality values form a narrow time region, and this time region seemingly corresponds to the beginning or the end of a time interval in which temporal edges occur in a bursty manner. In addition, by using TCC, we show that the remaining part of the temporal network is highly redundant in the sense that there are many ways to send information as quickly as possible. Although these properties are recognized in the network science community [28–30], we quantitatively confirm it for the first time using our centrality notions. We also demonstrate that the removal of temporal vertices according to their TBCC values is effective for hindering the propagation of information for both delaying and stopping it.

The paper is organized as follows. In Section 2, we introduce basic notions of temporal networks and the directed network associated with a temporal network. Section 3 introduces our centrality notions for temporal vertices, and Section 4 explains detailed methods of com-

puting our centrality notions. Section 5 is dedicated to demonstrating our experimental results. We give the conclusion in Section 6.

2 Preliminaries about temporal networks

2.1 Basic notions

We introduce the terminology and symbols to describe temporal network structure, which basically follow those used in Ref. [31].

For integer k, let $[k]$ denote the set $\{1, 2, \ldots, k\}$. We define $\mathbb{R}_+$ as the set of non-negative real numbers.

Let V be the set of vertices. A temporal edge is represented by quadruplet $e = (u, v, \tau, \lambda)$, where $u, v \in V$, $\tau \in \mathbb{R}$, and $\lambda \in \mathbb{R}_+$. For temporal edge $e = (u, v, \tau, \lambda)$, we refer to τ, λ, and $\tau + \lambda$ as the starting time, the duration, and the ending time of e, respectively. Temporal network $G = (V, E)$ is a pair of set of vertices V and set of temporal edges E.

When we study temporal networks, a vertex at a certain time is of interest. Therefore, we define a temporal vertex by a pair of vertex $v \in V$ and time $\tau \in \mathbb{R}$. In the following, we always use bold symbols such as $\boldsymbol{v}$ to denote temporal vertices. For temporal vertex $\boldsymbol{v} = (v, \tau)$, we denote the time τ by $\tau(\boldsymbol{v})$.

Temporal path P in temporal network $G = (V, E)$ is defined as an alternating sequence of temporal vertices and edges $P = \langle \boldsymbol{v}_1, e_1, \boldsymbol{v}_2, e_2, \ldots, e_{k-1}, \boldsymbol{v}_k \rangle$ satisfying the following properties. Let $\boldsymbol{v}_i = (v_i, \tau_i)$ for each $i \in [k]$. Then for each $i \in [k-1]$, the i-th temporal edge e_i is of the form $e_i = (v_i, v_{i+1}, \tau, \lambda)$ such that $\tau_i \leq \tau$ and $\tau + \lambda \leq \tau_{i+1}$. We define the starting time, the duration, and the ending time of P as τ_1, $\tau_k - \tau_1$, and τ_k, respectively. For two temporal vertices $\boldsymbol{u}$ and $\boldsymbol{v}$, relationship $\boldsymbol{u} \rightsquigarrow \boldsymbol{v}$ indicates that there is a temporal path from $\boldsymbol{u}$ to $\boldsymbol{v}$.

We define the earliest arrival time at vertex w when departing from temporal vertex $\boldsymbol{v}$ by the smallest $\tau \in \mathbb{R}$ such that $\boldsymbol{v} \rightsquigarrow (w, \tau)$, and we denote it by $\tau_{\mathrm{eat}}(\boldsymbol{v}, w)$. If there is no such τ, we define $\tau_{\mathrm{eat}}(\boldsymbol{v}, w) = \infty$. Similarly, we define the latest departure time from a vertex u for arriving at $\boldsymbol{v}$ as the largest $\tau \in \mathbb{R}$ such that $(u, \tau) \rightsquigarrow \boldsymbol{v}$, and we denote it by $\tau_{\mathrm{ldt}}(\boldsymbol{v}, u)$. If there is no such τ, we define $\tau_{\mathrm{ldt}}(\boldsymbol{v}, u) = -\infty$. A fastest temporal path from temporal vertex $\boldsymbol{v}$ to vertex w is a temporal path from $\boldsymbol{v}$ to $(w, \tau_{\mathrm{eat}}(\boldsymbol{v}, w))$, and a fastest temporal path from a vertex u to a temporal vertex $\boldsymbol{v}$ is a temporal path from $(u, \tau_{\mathrm{ldt}}(\boldsymbol{v}, u))$ to $\boldsymbol{v}$.

2.2 Directed acyclic graph representation

A directed acyclic graph (DAG) is a directed network with no directed cycle. In this section, we describe the DAG representation of a temporal network, which is useful when solving problems related to temporal paths and describing the centrality notions we will introduce in Section 3. This DAG representation and its variants have been considered in the analysis of temporal networks [17, 32–36].

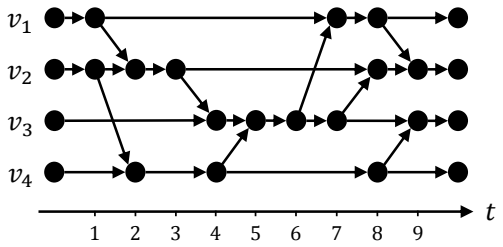

Fig. 2. DAG representation of the temporal network shown in Fig. 1.

For temporal network $G = (V, E)$, the DAG representation of G, denoted by $\widehat{G} = (\widehat{V}, \widehat{E})$, is constructed as follows. A vertex in $\widehat{G}$ represents a temporal vertex in G. For each $v \in V$, we first add to $\widehat{V}$ two vertices corresponding to the temporal vertices $(v, -\infty)$ and (v, ∞). For each temporal edge $(u, v, \tau, \lambda) \in E$, we add to $\widehat{V}$ two vertices corresponding to temporal vertices $\boldsymbol{u} = (u, \tau)$ and $\boldsymbol{v} = (v, \tau + \lambda)$ (if they do not exist in $\widehat{V}$) and add edge $(\boldsymbol{u}, \boldsymbol{v})$ to $\widehat{E}$. Finally, for each pair of temporal vertices $\boldsymbol{u} = (u, \tau), \boldsymbol{u}' = (u, \tau')$ sharing the same vertex u, we add edge $(\boldsymbol{u}, \boldsymbol{u}')$ to $\widehat{E}$ if there is no temporal vertex of the form (u, τ'') in $\widehat{V}$ such that $\tau < \tau'' < \tau'$.

Figure 2 illustrates DAG representation $\widehat{G}$ of temporal network G shown in Fig. 1. The vertex in the i-th row and the j-th column corresponds to the temporal vertex (v_i, j). For example, since there is temporal edge $(v_1, v_2, 1, 1)$ in G, we have an edge from $(v_1, 1)$ to $(v_2, 2)$ in $\widehat{G}$. For the i-th row, the leftmost and rightmost vertices correspond to the temporal vertices $(v_i, -\infty)$ and (v_i, ∞), respectively.

From the construction of the DAG representation, we have the following useful properties:

Lemma 1 *Let G be a temporal network. Then, $\widehat{G}$ is a DAG.*

Proof This is clear as we only add edges of the form $((u, \tau), (v, \tau'))$, where $\tau < \tau'$.

Lemma 2 *Let G be a temporal network. Suppose that temporal vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ have corresponding vertices in $\widehat{G}$. Then, there is a temporal path from $\boldsymbol{u}$ to $\boldsymbol{v}$ in G if and only if there is a directed path from $\boldsymbol{u}$ to $\boldsymbol{v}$ in $\widehat{G}$.*

Proof Let $P = \langle \boldsymbol{v}_1, e_1, \boldsymbol{v}_2, \ldots, e_{k-1}, \boldsymbol{v}_k \rangle$ be a temporal path from $\boldsymbol{v}_1 = \boldsymbol{u}$ to $\boldsymbol{v}_k = \boldsymbol{v}$. Without loss of generality, we assume that the time of $\boldsymbol{v}_i$ is equal to the starting time of e_i or the ending time of $\boldsymbol{v}_{i-1}$. Then, each $\boldsymbol{v}_i$ has a corresponding vertex in $\widehat{G}$. Let $\boldsymbol{v}_i = (v_i, \tau_i^v)$ for each $i \in [k]$ and $e_i = (v_i, v_{i+1}, \tau_i^e, \lambda_i^e)$ for each $i \in [k-1]$. Then, there is a directed path $(v_1, \tau_1^v), (v_1, \tau_1^e), (v_2, \tau_1^e + \lambda_1^e), (v_2, \tau_2^v), (v_2, \tau_1^e), (v_2, \tau_2^e + \lambda_2^e), \ldots, (v_k, \tau_k^v)$ in $\widehat{G}$. The converse easily follows the correspondence explained above.

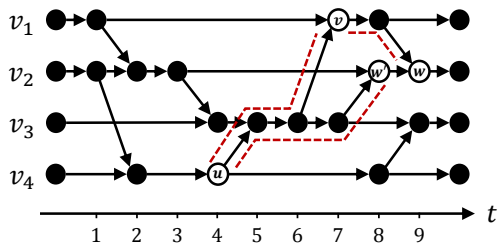

Fig. 3. Schematic describing the concept of temporal coverage centrality. The dashed polygonal lines represent the two temporal paths from vertex v_4 to v_2 that contain temporal vertex $\boldsymbol{v}$ in their durations.

3 Temporal coverage centralities

In this section, we introduce the temporal coverage centrality and the temporal boundary coverage centrality.

3.1 Temporal coverage centrality

Before defining TCC, we define the notion of coverage in temporal networks by generalizing its original version in static networks [37] as follows. Let $\boldsymbol{v}$ be a temporal vertex and u, w be vertices. Let $\boldsymbol{u} = (u, \tau_{\rm ldt}(\boldsymbol{v}, u))$ and $\boldsymbol{w} = (w, \tau_{\rm eat}(\boldsymbol{v}, w))$. Then, we say that $\boldsymbol{v}$ covers node pair (u, w) if the following two conditions hold:

1. $\tau_{\rm eat}(\boldsymbol{u}, w) = \tau_{\rm eat}(\boldsymbol{v}, w)$,
2. $\tau_{\rm ldt}(\boldsymbol{w}, u) = \tau_{\rm ldt}(\boldsymbol{v}, u)$.

In words, the earliest arrival time at w when departing from $\boldsymbol{u}$ does not change even if we drop by $\boldsymbol{v}$ (condition 1), and the latest departure time from u for arriving at $\boldsymbol{w}$ does not change even if we drop by $\boldsymbol{v}$ (condition 2). Figure 3 explains condition 1. Let us focus on $\boldsymbol{v} = (v_1, 7)$. Then, temporal vertices $\boldsymbol{u} = (v_4, \tau_{\rm ldt}(\boldsymbol{v}, v_4)) = (v_4, 4)$ and $\boldsymbol{w} = (v_2, \tau_{\rm eat}(\boldsymbol{v}, v_2)) = (v_2, 9)$ are determined as shown in the figure. We observe that, if we depart from $\boldsymbol{u}$ and are not forced to drop by $\boldsymbol{v}$, we can arrive at $\boldsymbol{w}' = (v_2, 8)$, which is earlier than $\boldsymbol{w}$. Hence, node pair (u, w) is not covered by $\boldsymbol{v}$ but by $\boldsymbol{w}'$.

On the basis of this notion of coverage, the TCC value of $\boldsymbol{v}$ is defined as the fraction of pairs $(u, w) \in V \times V$ that are covered by $\boldsymbol{v}$. By definition, the TCC value of a temporal vertex takes a real number in $[0, 1]$. If the TCC value is close to unity, the temporal vertex is said to be central in the sense that it covers many pairs of nodes. The formal definition is given in Algorithm 1 in an algorithmic manner.

3.2 Temporal boundary coverage centrality

Let $\boldsymbol{v} = (v, \tau)$ be a temporal vertex and u, w be vertices. Let $\boldsymbol{u} = (u, \tau_{\rm ldt}(\boldsymbol{v}, u))$ and $\boldsymbol{w} = (w, \tau_{\rm eat}(\boldsymbol{v}, w))$. Even if the TCC value of $\boldsymbol{v}$ is large, it does not always imply that removing the temporal edges involving $\boldsymbol{v}$ makes $\tau_{\rm eat}(\boldsymbol{u}, w)$ larger or $\tau_{\rm ldt}(\boldsymbol{w}, u)$ smaller. One particular reason for this is that sometimes we can reach v from $\boldsymbol{u}$ earlier than τ and can leave v later than τ to reach $\boldsymbol{w}$ (see temporal vertices $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ in Fig. 4). In some applications, we may want to regard such $\boldsymbol{v}$ as unimportant.

Algorithm 1 (The TCC value of $\boldsymbol{v}$)

$r \leftarrow 0$.
for $u \in V$ and $w \in V$ **do**
  $\boldsymbol{u} \leftarrow (u, \tau_{\rm ldt}(\boldsymbol{v}, u))$.
  $\boldsymbol{w} \leftarrow (w, \tau_{\rm eat}(\boldsymbol{v}, w))$.
  if $\tau_{\rm eat}(\boldsymbol{u}, w) = \tau(\boldsymbol{w})$ and $\tau_{\rm ldt}(\boldsymbol{w}, u) = \tau(\boldsymbol{u})$ **then**
    $r \leftarrow r + 1$.
return $r/|V|^2$.

To address this issue, we define TBCC by imposing additional criteria to the notion of coverage as follows. Note that, if focal temporal vertex $\boldsymbol{v}$ is an example of the situation stated in the previous paragraph, then $\tau_{\rm eat}(\boldsymbol{u}, v) < \tau$ or $\tau_{\rm ldt}(\boldsymbol{w}, v) > \tau$ should hold. Hence, we define that a pair (u, w) of vertices is covered at a boundary by temporal vertex $\boldsymbol{v}$ if the following hold:

1. (u, w) is covered by $\boldsymbol{v}$, and
2. $\tau_{\rm eat}(\boldsymbol{u}, v) = \tau$ or $\tau_{\rm ldt}(\boldsymbol{w}, v) = \tau$.

We explain this definition using the example shown in Fig. 4. Let $\boldsymbol{v}_i = (v, \tau_i)$ for $i \in [4]$. Note that all $\boldsymbol{v}_i$ ($i \in [4]$) cover vertex pair (u, w) as $\boldsymbol{u} = (u, \tau_{\rm ldt}(\boldsymbol{v}_i, u))$ and $\boldsymbol{w} = (w, \tau_{\rm eat}(\boldsymbol{v}_i, w))$ hold for all $i \in [4]$. In addition, note that all $\boldsymbol{v}_i$ cover (u, w). We can see that $\boldsymbol{v}_1$ and $\boldsymbol{v}_4$ cover (u, w) at the boundary because $\tau_{\rm eat}(\boldsymbol{u}, v) = \tau_1$ and $\tau_{\rm ldt}(\boldsymbol{w}, v) = \tau_4$. By contrast, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ do not cover (u, w) at the boundary.

On the basis of this notion of coverage at the boundary, the TBCC value of $\boldsymbol{v}$ is defined as the fraction of pairs (u, w) that are covered at the boundary by $\boldsymbol{v}$. Similar to TCC, the TBCC value of a temporal vertex takes a real number in $[0, 1]$ by definition. The formal definition is given in Algorithm 2 in an algorithmic manner.

In closing this section, it should be noted the difference between the previous notion of the temporal betweenness centrality and TCC (and TBCC). The main difference lies in the normalization of the number of vertex pairs covered by the temporal vertex. The definitions of TCC and TBCC do not normalize the number of such vertex pairs with the number of the fastest temporal paths, whereas the previous temporal betweenness centrality divides the number of the fastest paths that use the focal temporal vertex by the total number of the fastest temporal paths in the focal time window, as the betweenness centrality for static networks does [15–18]. We took such definitions of TCC and TBCC for the following reasons. First, TCC and TBCC become free from any parameters because we do not need to set the time window to count the number of the relevant fastest temporal paths for the normalization. Second, the TCC and TBCC values are easy to interpret as the fraction of the vertex pairs that have a fastest temporal path using the focal temporal vertex.

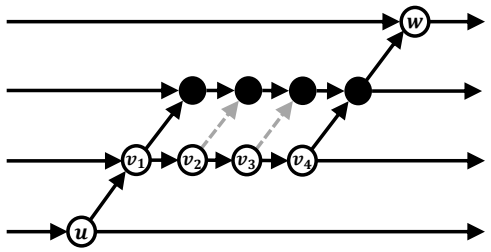

Fig. 4. Schematic describing the concept of temporal boundary coverage centrality. The dashed arrows represent the temporal edges that do not contribute the centrality values of the source temporal vertices.

Algorithm 2 (The TBCC value of $\boldsymbol{v}$)

```
r ← 0.
for u ∈ V and w ∈ V do
    u ← (u, τ_ldt(v, u)).
    w ← (w, τ_eat(v, w)).
    if τ_eat(u, w) = τ(w) and τ_ldt(w, u) = τ(u) then
        if τ_eat(u, v) = τ(v) or τ_ldt(w, v) = τ(v) then
            r ← r + 1.
return r/|V|^2.
```

4 Computing temporal coverage centralities

We can straightforwardly calculate TCC and TBCC according to Algorithms 1 and 2. In this section, to manage large temporal networks, we give efficient methods for computing TCC and TBCC on the basis of a graph–indexing technique developed recently in the database community [27], in particular, the method proposed in [42]. The key idea is in how to speed up the computation of $\tau_{\rm eat}$ and $\tau_{\rm ldt}$ in Algorithms 1 and 2. We describe the exact computation of TCC and TBCC in this section, and we also give the algorithms to approximate the TCC and TBCC values whose running time is polylogarithmic in the total number of vertices in G (see Appendix B).

In a directed network, we say that a vertex v_t is reachable from v_s if there is a directed path from v_s to v_t. With respect to Lemma 2, to enumerate the number of pairs (u, w) being covered by $\boldsymbol{v}$ (at the boundary, if needed), we want to efficiently answer reachability in the DAG representation $\widehat{G}$ of given temporal network G. To this end, it is beneficial to construct an index of $\widehat{G}$ that computes the reachability between any pair of nodes on the basis of information of the reachability between a fraction of node pairs. Such an index is often called a reachability oracle in the database community [38–42].

The basic idea of the construction of a reachability oracle for the present problem is the following. Naively, we want to compute a large table that stores the reachability of every pair of temporal vertices. If this were possible, we could answer reachability just by looking at that table. Unfortunately, however, perfecting this table requires $O(|\widehat{V}|^2)$ computation time and $O(|\widehat{V}|^2)$ space, which could be prohibitively slow and large. The reachability oracle overcomes this problem by carefully storing partial information of the network. Based on the information, it efficiently computes the reachability for the whole network.

The method proposed in Ref. [42], which we will use for the numerical experiments in Section 5, computes a small table for each temporal vertex that stores reachability from (and to) a smaller number of other certain temporal vertices than the number of all the temporal vertices. It depends on the structure of each temporal network how small the table becomes. Then, we can answer the reachability from a temporal vertex $\boldsymbol{u}$ to a temporal vertex $\boldsymbol{v}$ by checking whether there is another temporal vertex $\boldsymbol{w}$ such that we can confirm the reachability from $\boldsymbol{u}$ to $\boldsymbol{w}$ and from $\boldsymbol{w}$ to $\boldsymbol{v}$ using the small tables of $\boldsymbol{u}$ and $\boldsymbol{v}$. If there is such $\boldsymbol{w}$, we indeed have a directed path from $\boldsymbol{u}$ to $\boldsymbol{v}$. The challenging part of the construction lies in guaranteeing the other direction; if there is a directed path from $\boldsymbol{u}$ to $\boldsymbol{v}$, then there is always such $\boldsymbol{w}$. In addition, we need to be able to compute the small table for each vertex efficiently. This method resolves these issues, so that it can handle directed networks of millions of edges with the query time of less than a microsecond on average (see Ref. [42] for further technical details).

Table 1. Basic statistics of the datasets. Variables n, m, $\widehat{n}$, and $\tau_{\max}$ are the total number of vertices and temporal edges in G, the total number of vertices in $\widehat{G}$, and the maximum ending time of a temporal edge, respectively. The datasets are arranged in increasing order of m.

Name	n	m	$\widehat{n}$	$\tau_{\max}$
Infectious [43]	410	17298	32218	1393
HT09 [43]	113	20187	48477	5246
Hospital [44]	75	32424	65296	9454
Irvine [45]	1899	59835	220772	58192
Email [46]	167	82927	254533	57843

5 Results

The basic statistics of the datasets we use are summarized in Table 1. It should be noted that we do not use the actual time stamps in the datasets but define τ by the order of unique values of the time stamps. For example, if the dataset consists of two time stamps $t = 1, 4$, we translate them into $\tau = 1, 2$. In addition, we assume that λ is equal to the finest time resolution of each dataset for all the temporal edges. Although interactions in Irvine and Email are directed (i.e., from sender to receiver(s) of messages), we regard them as undirected.

5.1 Statistics of TCC and TBCC

Figure 5 depicts the rank plots of the TCC and TBCC values of temporal vertices in the decreasing order. In all the datasets except for the Email data, at least 10% of temporal vertices have TCC values larger than 0.1 (Fig. 5(a)). This fact implies the redundancy of temporal networks in the sense that, when information flows between temporal vertices, it can drop by different vertices without

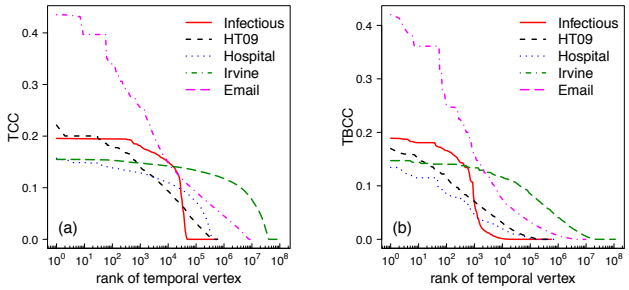

Fig. 5. Rank plots of the (a) TCC and (b) TBCC values.

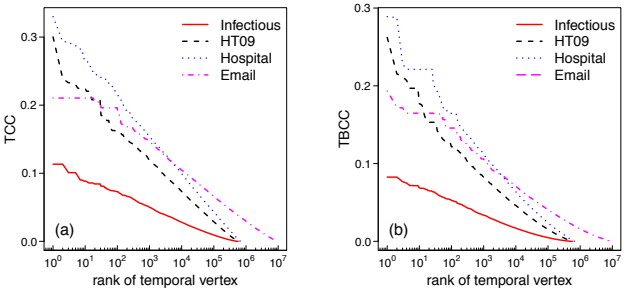

Fig. 6. Rank plots of the (a) TCC and (b) TBCC values in randomized temporal networks. The curves for Irvine are not provided because the computation did not stop.

increasing the total duration of the temporal paths. However, there are a smaller number of temporal vertices with large TBCC values (Fig. 5(b)). This fact also implies the redundancy of temporal networks in a different sense such that, when information flows between temporal vertices, it is not forced to exist at a certain vertex at a certain time.

To see the impact of the structural peculiarity of temporal networks on these distributions, we computed the centrality values of temporal vertices in randomized temporal networks. We randomize an original temporal network by replacing the two ends of each temporal edge by vertices chosen uniformly at random (similar to the procedure called randomized edges with randomly permuted times in Ref. [5]). The resultant centrality values are shown in Fig. 6. We notice that more temporal vertices have sufficiently large centrality values (e.g., larger than 0.1) in real-world temporal networks (Fig. 5) than in randomized temporal networks (Fig. 6). The maximum centrality values are larger in the randomized than in the original networks for HT09 and Hospital, and vice versa for Infectious and Email. This fact implies that the way the flow concentrates upon temporal vertices depends on each dataset.

It should be noted that the calculation for the randomized Irvine dataset did not stop even though the Email dataset, which has larger $\hat{n}$ than the Irvine, stopped. We can explain this result with the increase in the number of vertex pairs connected via temporal paths. The dominant factor of the computational time is the number of vertex connected via temporal paths because we have to consider all of such vertex pairs to calculate the centrality value of a temporal vertex. After the randomization, most of the vertex pairs are likely to have temporal paths and the number of such pairs scales with n^2. If we take into account that the Irvine dataset has the largest n value among the five datasets we consider, it makes sense for the Irvine dataset to require the far longer computational time compared to the other four datasets.

Next, we examine how the centrality values change over time owing to the structural transformation of the temporal networks. Figure 7 depicts the change in the maximum TCC and TBCC values over temporal vertices at present and the number of temporal vertices at present for Infectious and Hospital. In both datasets shown in Fig. 7, we can see some periodic patterns in the number of temporal vertices. However, the maximum centrality values are not much affected by the patterns, which implies that these values are determined not by the mere activity level in the networks but by the structure of the temporal network. In addition, the fact that the maximum centrality values vary considerably throughout the observation periods suggests that we should carefully incorporate temporal structure to assess the importance of vertices. Generally, the maximum TCC values are larger than the maximum TBCC values, which makes sense according to their definitions (i.e., TBCC only counts the coverage of the temporal paths at the boundary but TCC does not impose this boundary criterion).

When we focus on a particular vertex, two centrality values of it also vary in a different manner over time. Figure 8 depicts the change in the TCC and TBCC values of the vertex that are involved in the largest number of temporal edges in the two datasets, Infectious and Hospital. The TCC value of the vertex increases with time in Infectious (Fig. 8(a)), simply because the number of present temporal vertices increases and thus the focal vertex can reach these vertices in this period (also see Fig. 7(a)). By contrast, the TBCC value does not exhibit such an increasing trend. This fact supports our original purpose of introducing TBCC, i.e., to discount the centrality values of the temporal vertices of the dispensable temporal paths. In addition, the plot of TBCC unveils that even the vertex with the largest number of temporal edges does not always bridge effective temporal paths. In Hospital (Fig. 8(b)), we can observe that the temporal edges associated with the focal vertex are partitioned into five time intervals, in each of which temporal edges occur in a bursty manner, and the centrality values of the vertex become larger at the beginning and the end of each of these time intervals. This observation makes sense because, at the endpoints of a time interval, a vertex tends to play the role as the gateway for information flowing into or out of the time interval.

The computational efficiency of the two centralities enables us to draw a map of the centrality values of all the temporal vertices over time. This map reveals the existence of bottleneck time regions in the empirical temporal

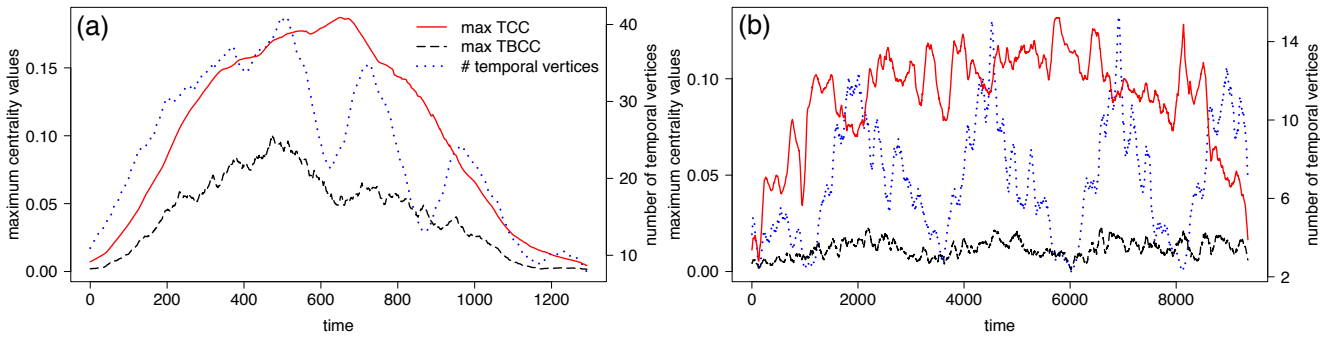

Fig. 7. Change in the maximum TCC and TBCC values over temporal vertices at present in (a) Infectious and (b) Hospital. For readability, we smoothed the curves by taking the average over a sliding window with a length of 100 units of time.

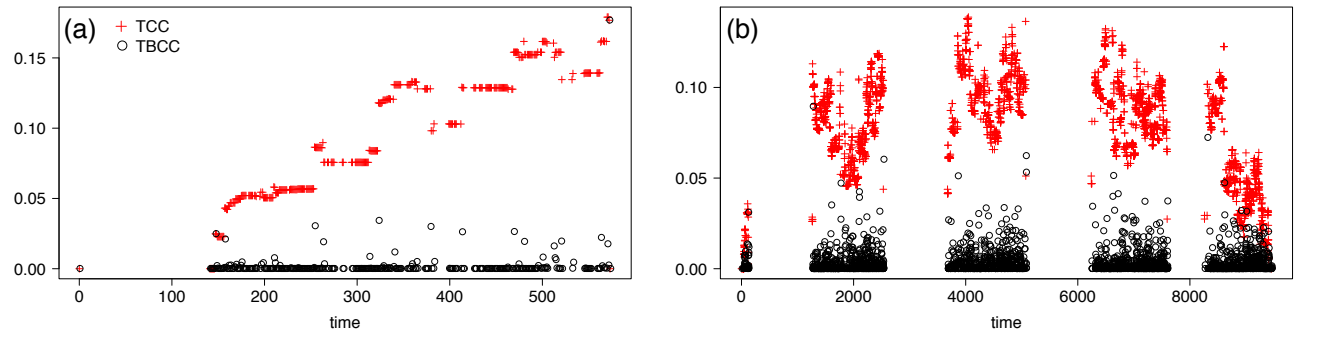

Fig. 8. Change in the TCC and TBCC values of the vertex with the largest number of temporal edges. (a) Vertex with label 195 in Infectious and (b) vertex with label 1115 in Hospital.

networks. Figures 9(a) and 9(b) depict the TCC values of temporal vertices as a heat map for Infectious and Hospital, respectively. In both datasets, most temporal vertices have non-negligible TCC values, and these results support the notion of redundancy of temporal networks (see Fig. 5(a)) such that all the vertices can belong to redundant temporal paths. In addition, the temporal vertices with the largest centrality values appear in the middle of the observation period (around time 700 and 6000 in Figs. 9(a) and (b), respectively), and the temporal vertices at the same time tend to have similar TCC values. We found the same phenomenon in all the datasets (see Electronic Supplementary Materials for the plots of the other datasets), and the existence of this bottleneck time period seems to be a common property of empirical temporal networks.

If we are interested in when these bottleneck time periods begin and end, we can look at the heat map of the TBCC values. As an example, Fig. 9(c) magnifies a bottleneck time period in Infectious (Fig. 9(a)) in which we observe many temporal vertices with the largest TCC values. However, the boundary of the bottleneck period is not clear in the figure. Figure 9(d) shows the heat map of the TBCC values in the same area as shown in Fig. 9(c). As we observe, the TBCC values indicate the boundaries at $\tau \simeq 660$, 680, and 750. This boundary information should be meaningful, for example, when we narrow the candidates of the vertices to be vaccinated for epidemic spreading on temporal networks [47–49].

We finally stress again that it becomes possible to compute these statistics and analyze the structure of temporal networks in such detail because of the efficient computation of TCC and TBCC using the reachability oracle.

5.2 Delay caused by removing a central temporal vertex

In closing this section, to verify the relevance of the proposed centrality notions at the microscopic level, we briefly report that removing a temporal vertex with large TCC and TBCC values is effective in delaying the propagation of information.

Let $G = (V, E)$ be a temporal network, where $V = \{v_1, v_2, \ldots, v_n\}$. For a temporal vertex $\boldsymbol{v} = (v, \tau)$, let $\boldsymbol{v}_i = (v_i, \tau_{\mathrm{eat}}(\boldsymbol{v}, v_i))$ for each $i \in [n]$ and τ' be the (unique) time such that $\boldsymbol{v}$ has an edge to $\boldsymbol{v}' = (v, \tau')$. We say that $\boldsymbol{v}_i$ gets prolonged by removing $\boldsymbol{v}$ if $\tau_{\mathrm{eat}}(\boldsymbol{v}, v_i)$ becomes larger by removing edges incident to $\boldsymbol{v}$ (and we keep edge $(\boldsymbol{v}, \boldsymbol{v}')$). In a similar manner, we say that $\boldsymbol{v}_i$ becomes disconnected by removing $\boldsymbol{v}$ if we cannot reach $\boldsymbol{v}_i$ from $\boldsymbol{v}$ after removing edges incident to $\boldsymbol{v}$ (where, again, we keep edge $(\boldsymbol{v}, \boldsymbol{v}')$).

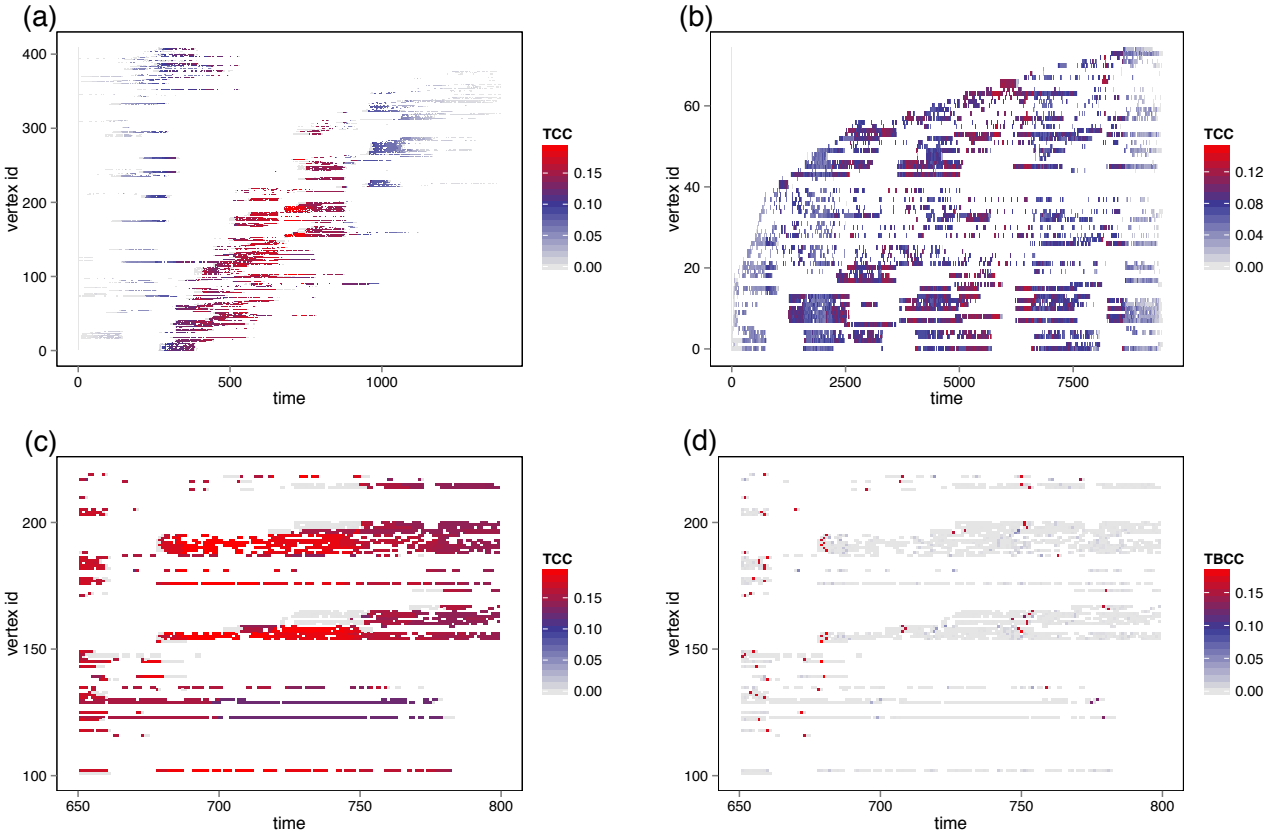

Fig. 9. Heat maps of the TCC values for (a) Infectious and (b) Hospital. (c) Heat map magnifying the area with $650 \leq \tau \leq 800$ and $100 \leq \mathrm{ID} \leq 220$ in (a). (d) Heat map of the TBCC values in the same area as shown in (c).

We investigate the fraction of prolonged or disconnected temporal vertices among $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$, by removing one of the top 100 vertices with respect to the TCC or TBCC values. It should be noted that the fraction of temporal vertices becoming prolonged or disconnected is nontrivial because the definition of TCC and TBCC take into account temporal paths both before and after the focal temporal vertex. As a baseline for comparison, we also conduct the same test by removing a temporal vertex chosen randomly. For the random case, we randomly choose 100 temporal vertices without replacement and take the average of the fraction of prolonged or disconnected temporal vertices for these 100 trials.

The results of the removal test of temporal vertices are summarized in Table 2 for the five datasets. As we expected, the removals according to the largest centrality values make more temporal vertices prolonged or disconnected than the random removals. The removals according to the largest TCC values tend to prolong a certain fraction of temporal vertices for all the datasets considered. However, it makes few temporal vertices disconnected. These outcomes make sense because the number of other temporal paths running alongside the temporal path going through the focal temporal vertex is not considered in TCC (also see Section 3.1). By contrast, the removals according to the largest TBCC values make a considerable fraction of temporal vertices prolonged and disconnected. Remarkably, 50.8% of the temporal vertices, on average, become disconnected from a removed temporal vertex in Irvine. There is no clear distinction between the results of the offline (i.e., Infectious, HT09, and Hospital) and online (i.e., Irvine and Email) networks.

6 Conclusions

We introduced two centrality notions for temporal networks—temporal coverage centrality and temporal boundary coverage centrality—to represent the importance of a temporal vertex by the fraction of vertex pairs that can or should use the temporal vertex when sending information as quickly as possible. Compared to centrality notions proposed in previous work, TCC and TBCC have two advantages: (i) Parameters or time windows do not need to be set and (ii) computation time is reasonable.

Applying TCC and TBCC to multiple datasets of empirical temporal networks, we revealed that there tends to be particular bottleneck time periods that play a crucial role in propagating information quickly and that the rest of the networks is redundant in the sense that there are many temporal paths to send information with the same duration. Although such structural redundancy in temporal networks was suggested in some previous studies [28–30], our centrality notions enable us to clearly quantify

Table 2. Results of the removal of temporal vertices. The number in each cell presents the average fraction of disconnected (or prolonged) temporal vertices over the 100 trials of the removal based on the given procedure (i.e., according to the largest TCC and TBCC values or random pick).

Dataset	TCC		TBCC		Random	
	Prolonged	Disconnected	Prolonged	Disconnected	Prolonged	Disconnected
Infectious	0.013	0.001	0.014	0.232	0.010	0.001
HT09	0.082	0.001	0.264	0.069	0.031	0.007
Hospital	0.049	0.001	0.156	0.257	0.037	0.001
Irvine	0.014	0.003	0.006	0.508	0.018	0.012
Email	0.136	0.006	0.375	0.016	0.054	0.000

and visualize this property. We believe that the centrality notions we proposed are useful for further studying the structure of temporal networks and verifying generative models of temporal networks.

Datasets used in the numerical experiments, Infectious, HT09, and Hospital were originally collected and published by the SocioPatterns collaboration (`http://www.sociopatterns.org/`). Datasets HT09 and Hospital were downloaded from the SocioPatterns website. Datasets Infectious, Irvine, and Email were downloaded from the Koblenz Network Collection (`http://konect.uni-koblenz.de/`). The authors thank Dr. James Cheng for valuable discussions. Yuichi Yoshida is supported by JSPS Grant-in-Aid for Young Scientists (B) (No. 26730009), MEXT Grant-in-Aid for Scientific Research on Innovative Areas (24106003), and JST, ERATO, Kawarabayashi Large Graph Project. T.T., Y. Yano and Y. Yoshida designed the research. Y. Yoshida constructed the algorithms to compute the centralities and gave the proof of their computational complexity. Y. Yano implemented the algorithms. T.T. analyzed the data sets. Y. Yano performed the numerical experiments of the removal of temporal vertices. T.T., Y. Yano, and Y. Yoshida discussed all the results and wrote the manuscript.

A Computational complexity of calculating τ_{eat} and τ_{ldt} with the reachability oracle

With the aid of the reachability oracle, we can efficiently compute τ_{eat} and τ_{ldt}:

Lemma 3 *Let G be a temporal network and $\widehat{G}$ be its DAG representation. We can compute τ_{eat} and τ_{ldt} with $O(\log|E|)$ queries to the reachability oracle of $\widehat{G}$.*

Proof We only consider τ_{eat} as τ_{ldt} can be computed similarly. Given temporal vertex $\boldsymbol{v}$ and vertex w, $\tau_{\mathrm{eat}}(\boldsymbol{v}, w)$ is the minimum $\tau \in \mathbb{R}$ such that there is a temporal path from $\boldsymbol{v}$ to (w, τ). To find such τ, we perform a binary search using the reachability oracle. Since the number of possible values for τ is $O(|E|)$, the number of queries is $O(\log|E|)$.

Lemma 4 *Let G be a temporal network and $\widehat{G}$ be its DAG representation. For any temporal vertex $\boldsymbol{v}$, we can compute the TCC and TBCC values of $\boldsymbol{v}$ with $O(|V|^2 \log|E|)$ queries to the reachability oracle of $\widehat{G}$.*

Algorithm 3 (Approximation to the TCC value of $\boldsymbol{v}$)

```
r ← 0.
for i = 1 to k := (1/(2ε²)) log(2|V|²) do
    Sample vertices u, w ∈ V uniformly.
    u ← (u, τ_ldt(v, u)).
    w ← (w, τ_eat(v, w)).
    if τ_eat(u, w) = w and τ_ldt(w, u) = u then
        r ← r + 1.
return r/k.
```

Proof The proof is immediate from Lemma 3 and the algorithm definitions of TCC (Algorithm 1) and TBCC (Algorithm 2).

B Approximate computation of temporal coverage centralities

By Lemma 4 (see Section 4), the number of queries to the reachability oracle for computing the TCC and TBCC values is (almost) quadratic in the number of vertices of a temporal network. However, in some applications, we may want to compute these centralities faster. Here, we introduce a standard technique that enables us to approximate these centrality values with a sublinear number of queries. We only explain the case of TCC; the case of TBCC is performed in a similar way.

Algorithm 3 is an approximate method for computing the centrality value. The difference from Algorithm 1 is that, instead of enumerating all pairs (u, w), we only sample $O(1/\epsilon^2)$ pairs of vertices and take the average over them, where ϵ is the parameter controlling the possible error in approximation.

To show that Algorithm 3 gives a good approximation, we need to recall Hoeffding's inequality:

Lemma 5 (Hoeffding's inequality [50]) *Let $X_1, X_2, \ldots, X_k$ be independent random variables in $[0,1]$ and $\overline{X} = (1/k)\sum_{i=1}^{k} X_i$. Then, for any positive real number t,*

$$\Pr[|\overline{X} - \mathbf{E}[\overline{X}]| \geq t] \leq 2\exp(-2t^2 k).$$

Lemma 6 *Let G be a temporal network and $\widehat{G}$ be its DAG representation. For any temporal vertex $\boldsymbol{v}$, with $O(\log^2|V|/\epsilon^2)$ queries to the reachability oracle of $\widehat{G}$, we*

can compute the TCC value of $\boldsymbol{v}$ *with additive error of* ϵ *with probability of at least* $1 - 1/|V|^2$.

Proof Consider Algorithm 3 and let $\widetilde{C}(\boldsymbol{v})$ denote its output. Algorithm 3 issues $\mathrm{O}(\log^2 |V|/\epsilon^2)$ queries since τ_{ldt} and τ_{eat} can be computed with $\mathrm{O}(\log |V|)$ queries (see Lemma 3). Let X_i be the temporal edge at which we increment r in the i-th loop and $\overline{X} = (1/k)\sum_{i=1}^{k} X_i$. Note that $\mathbf{E}[\widetilde{C}(\boldsymbol{v})] = \mathbf{E}[\overline{X}] = (1/k)\sum_{i=1}^{k}\mathbf{E}[X_i] = C(\boldsymbol{v})$, where $C(\boldsymbol{v})$ is the TCC value of $\boldsymbol{v}$. Since $X_1, X_2, \ldots, X_k$ are independent random variables in $[0, 1]$, by Lemma 5, we have

$$\begin{aligned}
&\Pr[|\widetilde{C}(\boldsymbol{v}) - C(\boldsymbol{v})| \geq \epsilon] = \Pr[|\overline{X} - C(\boldsymbol{v})| \geq \epsilon] \\
&\leq 2\exp(-2\epsilon^2 \frac{1}{2\epsilon^2}\log(2|V|^2)) = 2\exp(-\log(2|V|^2)) \\
&= \frac{1}{|V|^2}.
\end{aligned}$$

Hence, the lemma holds.

Recalling that the query time of the reachability oracle is tiny, we find that the running time of Algorithms 3 can be seen as polylogarithmic in the input size. This is the great advantage of TCC and TBCC against other centrality notions.

References

1. R. Albert and A.-L. Barabási. *Rev. Mod. Phys.* **74**, 47 (2002).
2. M. E. J. Newman. *SIAM Rev.* **45**, 167 (2003).
3. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. *Phys. Rep.* **424**, 175 (2006).
4. M. E. J. Newman. *Networks: An Introduction.* (Oxford University Press, Oxford, UK, 2010).
5. P. Holme and J. Saramäki. *Phys. Rep.* **519**, 97 (2012).
6. P. Holme and J. Saramäki, (editors). *Temporal Networks.* (Springer-Verlag, Berlin, Heidelberg, 2013).
7. A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks.* (Cambridge University Press, Cambridge, MA, 2008).
8. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. *Rev. Mod. Phys.* **87**, 925 (2015).
9. D. Kempe, J. Kleinberg, and É. Tardos. In *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, 2003* (ACM Press, New York, 2003), p. 137.
10. A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747 (2004).
11. R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7794 (2005).
12. M. Kitsak, S. Havlin, G. Paul, M. Riccaboni, F. Pammolli, and H. E. Stanley. *Phys. Rev. E* **75**, 056115 (2007).
13. J. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész. *Phys. Rev. Lett.* **99**, 228701 (2007).
14. J. Tang, M. Musolesi, C. Mascolo, and V. Latora. In *Proceedings of the 2nd ACM workshop on Online social networks, Barcelona, Spain, 2009* (ACM Press, New York, 2009), p. 31.
15. J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora. *Phys. Rev. E* **81**, 055101 (2010).
16. J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia. In *Proceedings of the 3rd Workshop on Social Network Systems, Paris, France, 2010* (ACM Press, New York, 2010), no. 3.
17. H. Kim and R. Anderson. *Phys. Rev. E* **85**, 026107 (2012).
18. A. Alsayed and D. J. Higham. *Chaos, Solitons & Fractals* **72**, 35 (2015).
19. R. K. Pan and J. Saramäki. *Phys. Rev. E* **84**, 016105 (2011).
20. P. Grindrod, D. J. Higham, M. C. Parsons, and E. Estrada. *Phys. Rev. E* **83**, 046120 (2011).
21. P. Grindrod and D. J. Higham. *SIAM Rev.* **55**, 118 (2013).
22. E. Estrada, *Phys. Rev. E* **88**, 042811 (2013).
23. L. E. C. Rocha and N. Masuda. *New J. Phys.* **16**, 063023 (2014).
24. S. Motegi and N. Masuda. *Sci. Rep.* **2**, 904 (2012).
25. V. Nicosia, J. Tang, C. Mascolo, and M. Musolesi. In *Temporal Networks*, edited by P. Holme and J. Saramäki, (Springer-Verlag, Berlin, Heidelberg, 2013), p. 15.
26. D. A. Bader, S. Kintali, K. Madduri, and M. Mihail. In *Proceedings of the 5th international workshop on Algorithms and models for the web-graph, San Diego, CA, USA, 2007* (Springer-Verlag, Berlin, Heidelberg), p. 124.
27. J. X. Yu and J. Cheng. In *Managing and Mining Graph Data*, edited by C. C. Aggarwal and H. Wang, (Springer US, New York, 2010), p. 181.
28. S. Trajanovski, S. Scellato, and I. Leontiadis. *Phys. Rev. E* **85**, 066105 (2012).
29. T. Takaguchi, N. Sato, K. Yano, and N. Masuda. *New J. Phys.* **14**, 093003 (2012).
30. S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer. *IEEE Trans. Mobile Comput.* **12**, 105 (2013).
31. H. Wu, J. Cheng, S. Huang, Y. Ke, Y. Lu, and Y. Xu. *Proc. VLDB Endowment.* **7**, 721 (2014).
32. D. Kempe, J. Kleinberg, and A. Kumar. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing, Portland, OR, USA, 2000* (ACM Press, New York, New York, 2000), p. 504.
33. V. Kostakos. *Physica A* **388**, 1007 (2009).
34. R. Pfitzner, I. Scholtes, A. Garas, C. J. Tessone, and F. Schweitzer. *Phys. Rev. Lett.* **110**, 198701 (2013).
35. I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer. *Nat. Comm.* **5**, 5024 (2014).
36. L. Speidel, T. Takaguchi, and N. Masuda. *Eur. Phys. J. B* **88**, 203 (2015).
37. Y. Yoshida, In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2014* (ACM Press, New York, 2014), p. 1416.
38. K. Simon. *Theor. Comput. Sci.* **58**, 325 (1988).
39. E. Cohen, E. Halperin, H. Kaplan, and U. Zwick. *SIAM J. Comput.* **32**, 1338 (2003).
40. H. Yildirim, V. Chaoji, and M. J. Zaki. *Proc. VLDB Endowment* **3**, 276 (2010).
41. S. J. van Schaik and O. de Moor. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, Athens, Greece* (ACM Press, New York, 2011), p. 913.
42. Y. Yano, T. Akiba, Y. Iwata, and Y. Yoshida. In *Proceedings of the 22nd ACM international Conference on information and knowledge management, San Francisco, CA, USA, 2013* (ACM Press, New York, 2013), p. 1601.

43. L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck. *J. Theor. Biol.* **271**, 166 (2011).
44. P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B. Kim, B. Comte, and N. Voirin. *PLOS ONE* **8**, e73970 (2013).
45. T. Opsahl and P. Panzarasa. *Soc. Networks* **31**, 155 (2009).
46. R. Michalski, S. Palus, and P. Kazienko. In *Business Information Systems*, edited by W. Abramowicz, (Springer-Verlag, Berlin, Heidelberg, 2011), p. 197.
47. S. Lee, L. E. C. Rocha, F. Liljeros, and P. Holme. *PLOS ONE* **7**, e36439 (2012).
48. M. Starnini, A. Machens, C. Cattuto, A. Barrat, and R. Pastor-Satorras. *J. Theor. Biol.* **337**, 89 (2013).
49. N. Masuda and P. Holme. *F1000Prime Rep.* **5**, 6 (2013).
50. W. Hoeffding. *J. Am. Stat. Assoc.* **58**, 13 (1963).

Electronic Supplementary Material

for

Taro Takaguchi, Yosuke Yano, and Yuichi Yoshida

Coverage centralities for temporal networks

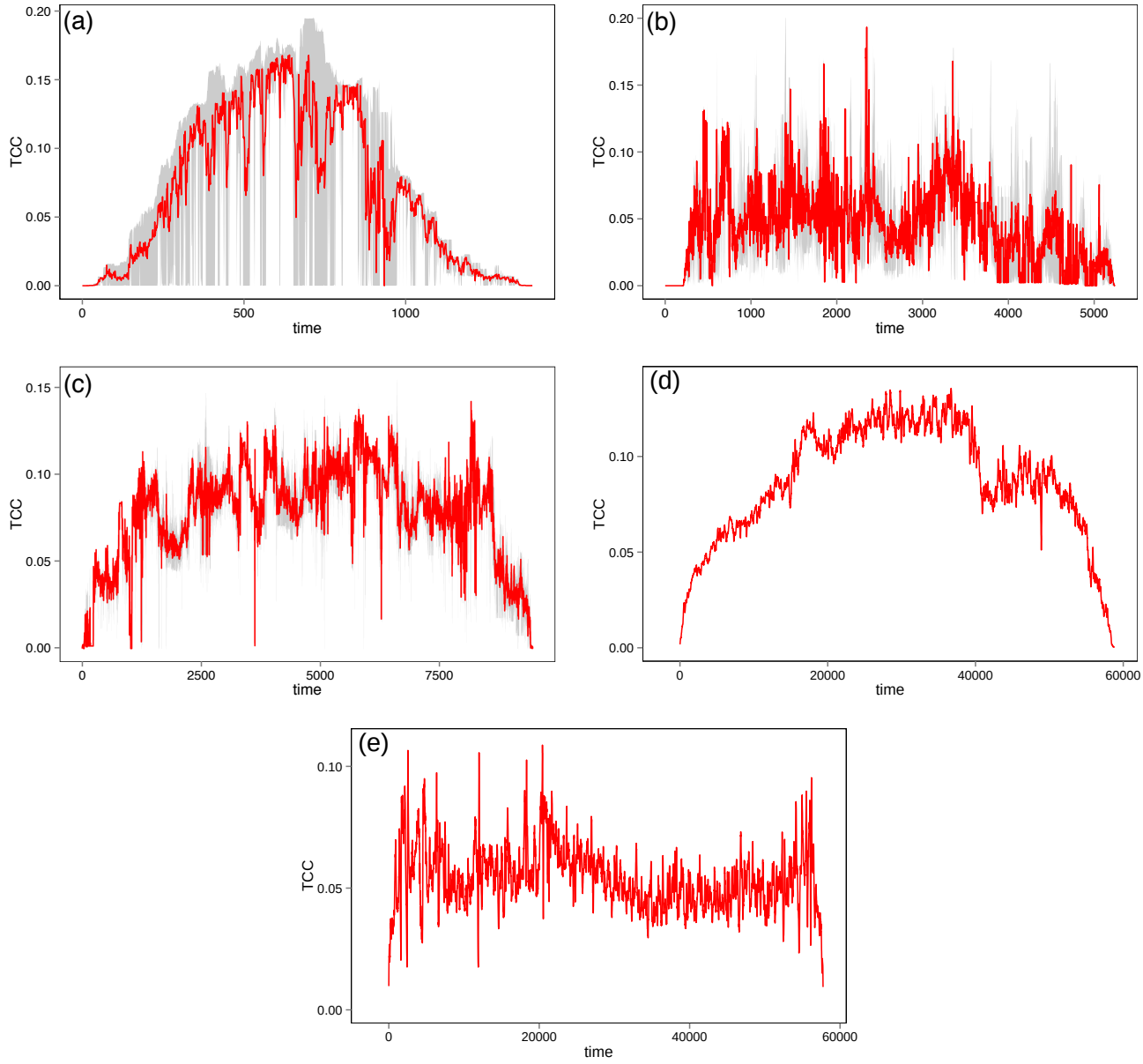

Fig. S1. Average (solid line) and $10-90\%$ values (shaded areas) of TCC at each time for (a) Infectious, (b) HT09, (c) Hospital, (d) Irvine, and (e) Email. We consider only the temporal vertices involved in temporal edges with other vertices to calculate the statistics. For (d) and (e), we smoothed the curves by taking the average over a sliding window with a length of 100 units of time, because the time resolutions of the observations are so high that there are not sufficient number of temporal vertices to take the average at most of the time points.